DOCUMENT RESUME

ED 279 686 TM 870 073

AUTHOR Lewis, Anne E.
TITLE The Politics of Testing/Assessment (Or a Chameleon in the Classroom).
PUB DATE 19 Aug 86
NOTE 23p.; One of 46 papers commissioned by the Study Group on the National Assessment of Student Achievement and cited in Appendix B to their final report "The Nation's Report Card" (TM 870 049). For other papers in this group, see TM 870 050-094.
PUB TYPE Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Educational Assessment; *Educational Testing; Elementary Secondary Education; Measurement Objectives; National Surveys; *Politics of Education; Public Opinion; Standardized Tests; State Surveys; *Testing Problems; *Testing Programs; *Test Results; *Test Use
IDENTIFIERS *National Assessment of Educational Progress

ABSTRACT

A number of political issues are involved in educational testing and the National Assessment of Educational Progress (NAEP). Many school officials are disturbed by the number of tests administered and the information and instructional time lost; they are reluctant to accede to further voluntary testing programs which might duplicate current efforts. State-by-state comparisons of published test results may reduce the amount of authority held by local school districts for decision making and curriculum formulation. Response to minimum competency testing, and the linking of teachers' merit pay or dismissal to student achievement have been strongly challenged. Groups' attitudes toward teaching to the test (i.e., coaching), and toward various uses of tests are quite subjective, but some groups are ready to do battle if assessment data are used in a punitive fashion. Other attitudes toward federal data collection efforts include: support for national testing if it is useful, desire for testing to not be duplicative or costly, and general trust in NAEP's program of a national assessment with state comparisons. Test scores tend to be reported simply, without explanations of their meaning; they should, however, be interpreted by education specialists. It is suggested that national testing be well-designed, avoiding compromises which would limit its usefulness. Proponents of school-based decision making and alternatives to standarized testing will counterbalance the support for national testing. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED279686

The Politics of Testing/Assessment

(Or a Chameleon in the Classroom)

Anne E. Lewis

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☒ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

A. C. Lewis

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper Commissioned by

THE STUDY GROUP ON THE NATIONAL ASSESSMENT OF STUDENT ACHIEVEMENT

1986

TM 810 073

BEST COPY AVAILABLE

THE POLITICS OF TESTING/ASSESSMENT

(Or a Chameleon in the Classroom)

Anne Lewis

August 19, 1986

doin.espacal

"The reasons for the problems surrounding assessment are political. People simply recognize that collecting test score data for policymaking purposes is a patently political activity, particularly when there isn't any scientific base for choosing the objectives for testing. Information can be and will be used as a political tool in the allocation of resources and in the allocation of jobs. Because of the stakes involved, people are going to fight over who is tested, what is tested, when it's tested, by whom, and how the data ought to be released."

Jerome Murphy, 1977

THE POLITICS OF TESTING/ASSESSMENT IN PUBLIC EDUCATION

(Or a Chameleon in the Classroom)

Testing/assessment of public education is a technical issue taken over by politicians. This is not a critical observation because, after all, the lady asked for it. Standardized testing is neither as mysterious and sacrosanct as psychometricians

believe nor as precise as politicians claim. Originally intended to help improve education programs, the use of standardized testing nationally in this country barely had begun to define its objectives before it became a tool for policymaking at state and federal levels.

This premature ascendency is creating some tough political problems. Pressures are mounting to use test results as the major factor in accountability, even though questions over the suitability of tests for policymaking purposes and the acceptance of such measures have yet to be answered. Yet, out of a volatile and undefined situation, a consensus among educators appears to be developing that will support state and national assessment efforts--within limits.

This paper will examine some of these political issues as candidly as is possible, considering the somewhat ephemeral environment in which they exist.

For a technique that began in France to identify intellectually subnormal children, standardized testing has led a chameleonic life. The progressives adopted it in the 1920-30s as a way to protect children from unfair judgments about children by teachers. At the same time, elite campuses began to use standardized assessment for admission decisions. Generally, its use by school districts remained a local option and issue, with critics now arguing that standardized testing had become an unfair way to judge students (and teachers).

The situation changed drastically about 10 years ago. The

dramatic decline in SAT scores triggered interest not only in what was happening in education but how it could be measured.

> "...For just about everyone with an intense interest in the public schools, the level of uncertainty surrounding the enterprise of education had become intolerable.... Legislators and governing boards viewed the use of test scores to assess learning as a way of freeing themselves from involvement in decisions related to program operation; at the same time, they knew that using test scores for this purpose would pressure educators to improve the schools and would produce evidence of improvement that the general public would find convincing." (Laura Salganik, Phi Delta Kappan, May, 1985)

About the same time standardized testing became a tool for mandated evaluations of federal programs. And if federal evaluators found such testing useful, why not apply it at state levels? A concurrent development was the increased centralization of education policymaking and funding at the state level, one that created demands for greater accountability from local school districts and resulted in the state-mandated minimal competency movement.

In fact, today there is an inevitable tie between any new or

increased investment in education by the states and numerical assessment of that effort--be it higher standards for students or higher salaries for teachers. As Scott Thomson, executive director of the National Association of Secondary School Principals, acknowledges, "We live in an outcome-based society. You can't buy a can of peas without a study and a consumer rating."

According to the Christian Science Monitor (Aug. 23, 1985), more than 100 test publishers provide $100 million worth of tests for K-12 education annually. Further, 40 states require minimum competency exams of students; 19 states require students to pass a test in order to graduate from high school (by 1987, there will be five more on that list); 32 states now require testing of those entering the teaching profession. Three states recently became embroiled in political turmoil because of mandates for testing of veteran teachers.

The conservative estimate is that 20 million standardized tests are given in K-12 education each year. At the other extreme, FairTest, a new watchdog organization, estimates that number to be 100 million. And the total testing industry--including higher education and professional exams--to be a $1 billion annual enterprise.

Supporters of these tests claim to be in the program improvement corner, although how much competency tests results, for example, are tied to better education of individual students is a very moot point. There is some crossover to the public

accountability corner, particularly with competency testing, but generally public information needs are supplied by other kinds of tests.

An exception was a recent report from the Congressional Budget Office, Trends in Educational Achievement. It notes rather emphatically--and goes at length to explain--that standardized test scores are not synonymous with educational achievement and that standardized tests have their limits. Then, it uses a half-dozen national standardized tests for its 165-page discussion of achievement trends.

Because the U.S. Department of Education began to use SAT and ACT scores as a comparative measure among states, these tests quickly became the basis for a "report card" to the public, despite unanimous sentiment from both their sponsors and educators that the tests were very inappropriate measures of school quality. When state-by-state SAT averages first appeared in a weekly education publication in 1981, the education community was horrified. But the coverage produced a heretofore publicly unknown fact--that the greater the number of students taking the SAT within a state, the lower the state score. Further, state comparisons opened up public discussion of the SAT in general, leading to policies in some states covering truth in testing and research confirming claims that coaching could improve SAT scores. Truth, allowed to compete in the open, wins out.

The role of the National Assessment of Educational Progress

also has expanded. Its cumulative information is much more useful as assessments begin to show trends over a decade or more--as far as a report to the public is concerned. Several states have demonstrated that NAEP is useful for gathering student achievement information within states, also, and taken the initial step of dealing with state-by-state comparisons. Virginia, Florida and Tennessee learned that the medicine was not so bitter, after all. NAEP's appropriateness and usefulness as a national indicator is, of course, a central political and policy question at this time.

Much less reliable as a measure of student achievement but perhaps one of the most popular resources cited by national reports and the press are international comparisons. A "homemade" test devised by a Dallas newspaper staff and given to a sample of that city's students as well as students in several countries--with disastrous results for the American students--no doubt contributed to the intense public debate in Texas over school reform. Protests from psychometricians and educators that the test was inaccurate and unfair couldn't match the public's concern over the results. Technical caveats about results from more formal testing, such as those conducted by the International Educational Achievement Organization (IEA), similarly have been overrun by the public's focus on what the test comparisons tell.

At this point, it is quite safe to say that the education community--teachers and administrators--accept the necessity of providing a public accounting of student achievement. Moreover,

state and federal policymakers are committed to collecting more sophisticated achievement data that will, among their design, provide state-by-state comparisons and measure more closely the effects of new state mandates and/or incentives for educational improvement.

This has happened in both a lightening and enlightening fashion. In the early 1970s, the National Education Association was so opposed to standardized testing that it called for a moratorium on such testing--a resolution which was dropped quietly 10 years later. Toward the end of the 1970s, American Federation of Teachers' President Albert Shanker began to speak out on the value of standardized testing results as a tool for public accountability, always using the occasion to distance AFT from the NEA position, which he called "extreme."

During this time, major administrator groups took little notice of any controversy in testing policies, other than to call for proper use of minimal competency results. Their lack of interest, compared to that of teachers, could have been related (and perhaps still is) to the fact that they control local testing policy decisions, while teachers, for the most part, do not.

The most remarkable change of attitude, however, was that of the state education chiefs. NAEP was allowed to happen only because the chiefs wrung an agreement that the data collected would be analyzed at no lower than a regional level. At the Council of Chief State School Officers annual meeting in Little

Council of Chief State School Officers annual meeting in Little Rock, Ark., in 1983, a few chiefs publicly broke ranks with this position, saying the time had come to accept state-by-state comparisons--and to take hold of the process themselves, rather than be hung every year on the U.S. Department of Education "wall chart."

A small group of chiefs studied the problem during the following year, and at the 1984 meeting, in Wilmington, Del., the chiefs adopted recommendations that would put them in charge of a national assessment effort allowing state-by-state comparisons. It was not an easy victory--won by proxy votes from those who probably were unaware of the importance of the vote. Nonetheless, once adopted, the idea generally was accepted. When refinements came up for a vote at the 1985 meeting, there was little debate.

Despite this refinement of attitudes by various groups, state and national testing efforts invite in several potentially risky political situations. They include:

- Get your act together
- Whose ox is being gored?
- Who is doing the goring?
- How does the score get reported

These political consequences will vary among the states, dependent more on political activity of certain groups than on the policies themselves.

Statements on the amount of time required for various types and levels of testing rely on anecdotal evidence. However, administrators particularly are beginning to argue that "enough is enough." The saturation issue cannot be dismissed as peripheral because if enough opposition to further or more complex testing programs emerges among teachers or administrators, a national testing program that relies on voluntary participation could be jeopardized.

Representative of the attitude of school district research directors are the remarks prepared for the Sounding Board of the Council of Chief State School Officers State Education Assessment Center by Walter Hathaway, director of research and evaluation for the Portland, Ore., public schools.

> "Many local school districts have carefully planned and developed student, school and district information systems. These systems have evolved over the last decade to more efficiently respond to the information needs of our teachers, students, parents, support staff, administrators, policymakers and community.... Embedded in them are multi-layered systems of student performance assessment carefully aligned with the curriculum our teachers are teaching and designed to help them teach it to all their students. These

information systems have earned the respect and trust of the constituencies they serve. We, therefore, have a 'show-me' attitude toward state and federal proposals to replace or add additional layers on top of these effective local information systems with others whose chief virtue is uniformity permitting cross-system comparisons. The cost of such replacements or additions is enormous in terms of the adjustments needed within even samples of over 15,000 school districts and the information and instructional time lost to our over 2.5 million teachers and other instructional staff and the 45 million students they serve."

The few responses that the Office of Educational Research and Improvement received in response to its proposal for a revamping of its statistical center frequently mentioned the need to prevent duplication of efforts. However, most of the responses were from individual chiefs, with a vested interest in preventing a parallel testing program from developing within the Statistics Center, as was proposed.

More telling are the comments of representatives of the major principal groups. "If I were a principal," says Ed Keller, deputy director of the National Association of Elementary Principals, "I wouldn't want two or three separate national testing programs hitting my school. That would be on top of

state tests, locally used national tests and locally developed tests. There's got to be a better way."

Secondary school principals, says Scott Thomson, executive secretary of the National Association of Secondary School Principals, would like one "sober, professionally done, well organized national testing program." The majority of principals, he says, believe in the monitoring of achievement, but the current national "report cards"--the annual Gallup Poll and SAT scores--are more frustrating than helpful. Principals have two "simple" requests about national assessment--that it be professional and that it not be redundant, he believes.

Whose Ox Is Being Gored?

If a national assessment effort limits itself primarily to a report card for the public--and leaves the follow-up to others--the political risks are minimal (provided, of course, the testing design is professional and the reporting is full and accurate). However, past experience makes some experts, many researchers and most student advocates skeptical. George Madeus of Boston College notes that "instead of using tests merely to inform policymakers, the results are now being employed by them to make things happen--automatically and mechanically--to individuals, to schools or to districts."

The loser in the linkage of state policy to testing has been local control, contends Salganik. "Governors and chief state

school officers," she wrote in the May 1985 Phi Delta Kappan, "seem certain to view public comparisons of average test scores across the states as justification for reasserting their own decision-making authority in education--an authority that they had earlier relinquished to local school districts. Indeed, governors and chief state school officers may support comparisons among the states primarily because such comparisons will give them an excuse to reduce local control."

Another phenomenon, spotted by Michael Kirst of Stanford University, is that some current reform efforts tie the education package together tightly, with textbook content, curriculum and testing much more under the control of state mandates than local initiative.

The flip side to control is to be so loose that the effort yields no effects. Competency testing is a good example. This phenomenon among the states has about run its course, and even its justification on the basis of improved public confidence is losing merit. The "minimal" word is getting to the public, and as passing rates reach high levels, the public sees this more as a fault of the tests than an indication of better performance by students. John Casteen III, while secretary of education for Virginia (which has a minimal competency exam) noted that "we have made more than our share of false starts. Minimum competency tests, for example, have rarely made a serious difference in our courses of instruction.... Students have passed or failed, but we have done little to build from this

year's experience toward next year's educational program."

However, where such tests are used as gatekeepers, the waters get much more political. Florida school officials were taken to court several times to prevent use of the state's literacy exam as a determination for high school graduation--stricly because of the racial performance on the test. Other states have been in the courts more recently over testing programs that limit access of minorities to teacher education programs. No research data are available yet, but some student advocacy groups already are linking higher dropout rates to reforms that rely heavily on test performance.

Merit pay plans that tie student test scores to teacher evaluation, likewise, are politically unpopular--although that hasn't stopped many school districts and some states from trying. The NEA has vociferously opposed this use of test scores; the AFT doesn't like it but hasn't officially said so. But it, too, has its limits. AFT recently filed suit in a federal court in St. Louis, Mo., against the school district's new policy of evaluating teachers solely on the basis of their students' scores on the California Achievement Test. This "up-or-out" policy went too far for the union, which wants an image of accommodation these days. The union's suit relies heavily on the research of Ronald Berk of The Johns Hopkins University, which found the use of test score data for merit pay eligibility--much less dismissal-- "indefensible."

Further, the issue of a national curriculum emanating from a

national assessment deserves attention. As presently reported, NAEP can only nudge. Its findings apparently have not influenced instruction or resources much, if at all. Were a national assessment to do more--even a more hardnosed approach to reporting the same amount of data--it could be viewed as leading to a national curriculum.

Yet, state testing apparently has resulted in a perception that there is widespread teaching to the test--with no organized opposition. Teacher leaders (NEA interviews) deplore it but say it's not their fault. Those from the principal groups and the Association for Supervision and Curriculum Development commented that if tests are properly designed and relate to curriculum goals, teaching to the test is not undesirable. (Those who really want to ride the bandwagon openly endorse the practice, under the rubric of "measurement-driven instruction".)

Generally, however, despite the acceptance of national assessment as a public relations tool, there are groups and special interests poised to do battle if assessment data are used punitively. Be helpful, says Thomson. But obviously not too helpful, indicate others.

Who Is Doing the Goring?

The "players" reacting to testing/assessment policies must depend more on what presently exists than on divining the effects of what possibly could be designed. And the results still are

quite subjective.

My interviews/research yielded three general opinions:

* As stated earlier, the education spokespeople accept the rationale for nationally collected data. However, they do not want duplication, they want the results to be useful and they do not want participation to be burdensome to local teachers and students. Further, it can be assumed that local educators would object if state resources substantially are redirected to national data gathering at the expense of providing data which they believe to be more relevant to local program improvement. Further, national data collection is perceived to be a federal role.

* However, the data collection proposed by OERI as a new Elementary and Secondary Integrated Data System--of student achievement in math and reading, school and teacher characteristics and family backgrounds--struck some who responded in writing as 1)duplicative of and competitive with new state efforts, and 2)potentially more of a policymaking tool than a data gathering one.

* Education groups tend to favor an expansion of NAEP as the vehicle for better data collection. One reason could be that NAEP findings have not meant much and, thus, are not perceived as threatening the status quo. Also, Gordon Cawelti, head of the Association for Supervision and Curriculum Development, pointed out that those who conceived NAEP expected it to provide information for policymaking by Congress--but Congress has never

used it. Another reason, most emphasized by Sharon Robinson, head of instructional services for NEA (but others, too), is that the education community trusts Greg Anrig, overseer of NAEP as president of the Educational Testing Service, to bring integrity and balance to testing policies. (I neglected to ask what the attitude would be if other leadership were in charge.)

How Do Scores Get Reported

If a national assessment is to be a report card, it has to be sent home. And if the past is any judge, the public will scan reporting based on a figure or two--and draw its inferences from that.

There is a chicken-or-the-egg aura to education reporting of tests/assessments. Editors--much more so than reporters--seem convinced that education reporting must be gussied up to appeal to readers or viewers, and it is true that education ranks low on readership appeal. Thus, there is a tendency to reduce test score reporting, for example, to a "pass/fail" simplicity or, when given choices of coverage that includes test scores, to report only scores. The "wall charts" have included as many footnotes to explain the data as data itself, but coverage of release of the wall charts ignore the explanations for the most part.

However, there is the possibility that readers are turned off by education coverage because it generally is dull. The education beat has a high turnover rate, few newspapers invest in education coverage and no major television network has an education specialist (except for a part-time consultant for the MacNeil/Lehrer Newshour on PBS).

Given background information and ideas for local application of national stories, reporters often can produce copy that provides depth to such issues as testing and appeals to readers. The Education Writers Association sponsored a seminar at the Wingspread Center last November on standardized testing. Clippings from that meeting showed that the information produced good local coverage--from a reporter's ability to provide background when his state issued "report cards" on its schools to a reporter's use of test experts to write about the bias issue in teacher testing.

However, these are the exception. The political realities must deal for a long time with national testing/assessment coverage that is superficial and skewed to single-figure measurements. And with the public perception that tests are a near-infallible way to evaluate education.

Given the current willingness (with some wariness) among education leaders to accept the idea of a national assessment with state comparisons, are there any contingencies? Yes, there are some considerations which need to be examined as possible

political problems for national assessment.

* The technical design and application of testing may fall short of assessment's promises. Mike Smith, a major author of the report leading to the CCSSO project on assessment, estimates that a proper assessment cannot be mounted without five years of planning. The most detailed response to OERI's proposal for a new data system criticized it for underestimating the sampling and other factors needed for an adequate assessment. If, as some commentators from education groups emphasized, the testing/assessment is not professionally, carefully designed, it will do more harm than good. A "sub-concern" would be if technical in-fighting jeopardized the support for a national assessment, such as differences over whether to use norm-referenced or criterion-referenced tests.

* A corollary of this is how to avoid compromises in a national assessment--both in objectives and in financial support--either one of which would limit its usefulness. It should be noted candidly that Congress does not seem inclined to provide additional funding for OERI.

* There are some counterbalancing forces opposed to reliance on national testing to drive educational change:

--If an assessment relies on the current mode of standardized tests, the Carnegie Foundation would be publicly

critical, with Ernest Boyer's objections now joined by those of Vito Perrone. The latter's work with the North Dakota Study Group, while limited to a few true believers, has kept alive the viability of alternatives to assessment based on tests. Upcoming reports by the foundation on the current reform effort will criticize a trend toward teaching to the test.

--FairTest, while a young advocacy organization, was very successful in promoting legislation in New York State this year to protect test takers. Its timetable calls for a campaign directed at elementary/secondary testing programs to begin within 18 months. Its goals are to assure bias-free testing and public scrutiny of testing.

--Alternatives to evaluation by testing are being promoted by those who can influence policy. Among them are Theodore Sizer of Brown University, who is working on different modes of evaluation through his Coalition of Essential Schools; and Project Zero at Harvard University, which is advocating student portfolios rather than standardized tests.

* The shift to school-based decisionmaking, advocated by the Carnegie report on teaching and the National Governors Association project 1991, among the latest, could reduce the emphasis on a national assessment. Local accountability could become the overrriding concern of the public. And in the future, the scenario might be quite different. Picture the reporting to a community on education progress in the hands of teachers rather

than principals, superintendents or research directors.

- New technologies or alternatives to the proposals being considered could influence the scope, expense and usefulness of a national assessment. For example, an option could be to merge existing state achievement data into an information base. This might be less intrusive than an expanded assessment, in the opinion of local educators. New technologies, on the other hand, which would link local data bases to state and national ones, might be seen as another step in the erosion of local decisionmaking.

A national assessment, attempting to provide state comparisons and to broaden the information the public receives about education, apparently must do much in a short time. If the public's interest in education is cyclical, as Kirst and others claim, this expanded assessment must begin to produce results probably before its technical integrity can match the political demands. It's risky.